

[illegible]

JOURNAL		elements in ciliated protozoa and a common "D35E" motif
MEDLINE	Pricz. Natl. Acad. Sci. U.S.A.	91 (3), 942-946 (1994)
REFERENCE	94134747	
AUTHORS	2 (bases 1 to 354)	
TITLE	Witherspoon,D.J., Doak,T.G., Williams,K., Seger,J. and Herrick,G.	
	Selection on the protein-coding genes of the TBEI family of transposable elements in the ciliates Oxytricha fallax and O. trifallax	
JOURNAL	unpublished	
REFERENCE	3 (bases 1 to 354)	
AUTHORS	Doak,T.G., Williams,K., Witherspoon,D.J. and Herrick,G.	
TITLE	Direct Submission	
JOURNAL	Submitted (11-FEB-1997) Oncological Science, University of Utah,	
	School of Med. RmSC334, USA, UT #A132, USA	
FEATURES	Location/Qualifiers	
source	. . . 354	
CDS	/organism="Oxytricha fallax"	
	/strain="gdl"	
	/transposon="TBEI"	
	/note="this is a bulk sequence that was generated from a PCR product that represents many transposon templates"	
	<1..>354	
	/codon_start=1	
	/product="57KD zinc finger/protein chimera"	
	/db_xref="PID:g1881576"	
	/translation=""HTFDLKKHLKAHKKXEXXXXXLKLRKAREXXXXXXXXXXQAXEXXXXXNNIXSEXTIKIMKIQQYKPVLQAQLDDDTLSQSYLLDESDRKVIXDHQYAKVPFNYTL"	
BASE COUNT	106 A 42 C 41 G 54 T 111 others	
ORIGIN		
Query Match	1.8%; Score 28; DB 111; Length 354;	
Best Local Similarity	30.1%; Pred. NO. 3.33e-03;	
Matches	25; Conservative 33; Mismatches 24; Indels 1; Gaps 1;	
Dd	110 hhyhdwargwbgydgwcgcyykaayaagcwamgarwyaswttrygltaargwdkrcrgca 169 ::: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: QY 802 CTGCTGAAGGTATGTAACCAATCAAACAAGAACCAGAA-TTAGTCAGAAGAATCATCCAA 860	
Dd	170 accwwyttrcmtcdgaagtmacd 192 ::::: : :: :: : QY 861 AGATAATTGCCTCGTGAAAA CA 883	
RESULT	4 standard: DNA; INV; 354 BP.	
ID	OXYRICHIA FALLAX	
AC	OXYTRICHA FALLAX	
N1	Eukaryotes; mitochondrial eukaryotes; Alveolata; Ciliophora; hypotrichs; Stichotrichida; Oxytrichidae; Oxytricha. [1] N1 1-354 MEDLINE; 94134747. RA Doak T.G., Doerder F.P., Jahn C.L., Herrick G.; RT "A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common 'D35E' motif"; Proc. Natl. Acad. Sci. U.S.A. 91:942-946(1994). [2] RN 1-354 RP Witherspoon D.J., Doak T.G., Williams K., Seger J., Herrick G.; RA "Selection on the protein-coding genes of the TBEI family of transposable elements in the ciliates Oxytricha fallax and O. trifallax"; RT unpublished. RL [3] RN 1-354 RP	

RA	Doak T.G., Williams K., Witherspoon D.J., Herrick G.:
RT	Submitted (11-FEB-1997) to the EMBL/GenBank/DBJ databases.
RL	Oncological Science, University of Utah, School of Med. Rm5C334,
RL	USA, UT 84132, USA
FH	Key Location/Qualifiers
FT	source
FT	1..354
FT	/organism="Oxytricha fallax"
FT	/strain="9D1"
FT	/transposon="TBE1"
FT	/note="this is a bulk sequence that was generated from a
FT	PCR product that represents many transposon templates"
FT	<1..>354
FT	/codon_start=1
FT	/product="57kd zinc finger/protein chimera"
FT	/transl_table=6
FT	/db_xref="PID:g1881676"
FT	/translation="HTRDLKHLKHLKAHKXXEXEXXXXLKXLLKKRKAREXXXXXXXXXX
..	
D	AXEXXXXXXXXXXXNXLSEXETKIMIKIYKKIPVLAQIDLDTSLSYLIEDSFDRKVIIX HOYAKVPENYLX"
FT	Sequence 354 BP; 106 A; 42 C; 41 G; 54 T; 111 other;
SO	
Query Match	1.8%; Score 28; DB 8; Length 354;
Best Local Similarity	30.1%; Pred. No. 3.33e-03;
Matches	Conservative 33; Mismatches 24; Indels 1; Gaps 1;
Db	110 hhyhdwargwbgydwcgcykaayagcagngarwyaswirrytargwdkrcgcga 169 ::: ::: ::: ::: ::: ::: :::
Oy	802 CTGGCGAAGTTATGGAACATCAACAACAAGCACAGA-TATGTCTAAGAGATCATCCA 860 ::: ::: ::: ::: ::: ::: :::
Db	170 accwyttrctmcddagrtmacd 192 ::: :::
Oy	861 AGATTATGACTCTCTGTGAAACA 883
RESULT	5 OFU89259 354 bp DNA INV 14-MAR-1997
LOCUS	Oxytricha fallax 57kd zinc finger/protein chimera gene, partial
DEFINITION	cds.
ACCESSION	U89259
NID	91881675
KEYWORDS	.
SOURCE	Oxytricha fallax.
ORGANISM	Eukaryota; mitochondrial eukaryotes; Alveolata; Ciliophora; Euhartzoae; Stichotrichida; Oxytricha. hypotrichs; 1 (bases 1 to 354) Doak,T.G., Doerder,F.P., Jahn,C.L. and Herrick,G. A proposed superfamily of transposable genes: transposon-like elements in ciliated protozoa and a common 'D35E' motif Proc. Natl. Acad. Sci. U.S.A. 91 (3), 942-946 (1994)
TITLE	2 (bases 1 to 354) Witherspoon,D.J., Doak,T.G., Williams,K., Seger,J. and Herrick,G. Selection on the protein-coding genes of the TBE1 family of transposable elements in the ciliates Oxytricha fallax and O. trifallax
JOURNAL	unpublished
REFERENCE	3 (bases 1 to 354)
AUTHORS	Doak,T.G., Williams,K., Witherspoon,D.J. and Herrick,G.
TITLE	Direct Submission
JOURNAL	Submitted (11-FEB-1997) Oncological Science, University of Utah, School of Med. Rm5C334, USA, UT 84132, USA
FEATURES	Location/Qualifiers
source	1..354 /organism="Oxytricha fallax" /strain="9D1" /transposon="TBE1" /note="This is a bulk sequence that was generated from a PCR product that represents many transposon templates"

[illegible]


```

/tissue_type="brain"
/cell_type="C6 glioma"
BASE COUNT      2124 a      1856 c      1799 g      1316 t
ORIGIN
Query Match      1.6%  Score 24:  DB 92;  Length 7095;
Best Local Similarity 76.1%;  Pred. No. 1.38e+00;
Matches 35;  Conservative 0;  Mismatches 11;  Indels 0;  Gaps 0;

Db      1004  gcaaggaatcagattcagcagcagtgagcagtggaaccaaca 1049
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY      305  GCAAGGAGCTCAGTACGTCAAGCAGAGTGCATCGACCCACAA 350

RESULT  10
LOCUS   PCHSNAB      7322 bp      DNA
DEFINITION P.sativum gene for chalcone synthase.
ACCESSION X80007
          g510542
ORIGINS
  peas.
  chalcone synthase; CHS-1A gene; CHS-1B gene; class 1 gene.
ORGANISM Pisum sativum
Eukaryotes; mitochondrial eukaryotes; Viridiplantae;
Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
Magnoliopsida; Rutanae; Sapindales; Fabaceae; Papilionoideae;
Pisum.
REFERENCE 1 (bases 1 to 7322)
AUTHORS Hellens,R.P.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 7322)
AUTHORS Hellens,R.P.
TITLE Direct Submission
SUBMITTED (01-JUL-1994) R.P. Hellens, John Innes Institute, Colney
Lane, Norwich NR4 7UH, UK
FEATURES
  source
    1..7322
    /organism="Pisum sativum"
    /strain="J1 813"
    /clone_lib="Iabda GEM-11"
    /clone="B"
    /map="linkage group 3"
    <1040..1217
    /gene="CHS-1A"
    /number=1
    join(<1040..1217,1328..>2319)
    /gene="CHS-1A"
    join(1040..1217,1328..2319)
    /gene="CHS-1A"
    /EC_number="2.3.1.74"
    /codon_start=1
    /product="naringenin-chalcone synthase"
    /db_xref="PID:g510543"
    /translation="MTVNEIRQARAGSPATVATGATPNCVCOSTYDPDFRIT
    NSHRTLKKEFORMCDKSKIRKRYMLTEILKENPSCEYAPSLDARDQVAVVEV
    PKLGEKATVAKLEWGPQKSKITLHICTTSGVDMPADQUTKLGLRIVYRVMY
    QOCFAGTALRLAKDLAENNGARVLVWSEITAVFRGSDTHLDSLVGALFCDG
    AAATVIGSDPLPOVEKPLFELWTAQTIIVPDSGALIDGRLREGTLFHLKQVPLIS
    KNIERAIYAEFQPLIGISDYNLSFWTAHGGPAILDOVEAKLGLKORKMOATRHVLSY
    GNMSSACVLFILDEMRKSKEDLATIGELGVLFGRPGLTVEIVLHSHAT"
    1218..1327
    /gene="CHS-1A"
    /number=1
    1328..>2319
    /gene="CHS-1A"
    /number=2
    join(<4336..4513,4642..>5633)
    /gene="CHS-1B"
    join(4336..4513,4642..5633)
    /gene="CHS-1B"
    /EC_number="2.3.1.74"
    /codon_start=1
    /product="naringenin-chalcone synthase"
    /db_xref="PID:g510544"

```

```

/translation="MTVDEIRQARAGSPATVLAIGTATPNCVDOSTYDPYFRIT
NSEHRTLEKFORMCDKSKIRKRYMLTEILKENPSVCETMAVSLDARDQVAVVEV
PRLGKEATVAKLEWGPQKSKITLHICTTSGVDMPADQUTKLGLRIVYRVMY
QOCFAGTALRLAKDLAENNGARVLVWSEITAVFRGSDTHLDSLVGALFCDG
AAATVIGSDPLPOVEKPLFELWTAQTIIVPDSGALIDGRLREGTLFHLKQVPLIS
KNIERAIYAEFQPLIGISDYNLSFWTAHGGPAILDOVEAKLGLKORKMOATRHVLSY
GNMSSACVLFILDEMRKSKEDLATIGELGVLFGRPGLTVEIVLHSHAT"
<4336..4513
/gene="CHS-1B"
/number=1
4514..4641
/gene="CHS-1B"
/number=1
4642..>5633
/gene="CHS-1B"
/number=2
BASE COUNT      2350 a      1307 c      1264 g      2401 t
ORIGIN
Query Match      1.6%  Score 24:  DB 64;  Length 7322;
Best Local Similarity 81.6%;  Pred. No. 1.38e+00;
Matches 31;  Conservative 0;  Mismatches 7;  Indels 0;  Gaps 0;

Db      654  aaagtgtaacctctatggaataaccaccaatggtt 691
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
OY      1404  AAAGACTAATCATCTCTCCATTAACCCCAATGCTT 1441

RESULT  11
LOCUS   CEC29F3      18632 bp      DNA
DEFINITION Caenorhabditis elegans cosmid C29F3.
ACCESSION Z81043
NID      91627634
KEYWORDS
SOURCE   Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryotes; mitochondrial eukaryotes; Metazoa; Nematoda;
Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;
Rhabditidae; Pelodierinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 18632)
AUTHORS Matthews,L.
JOURNAL Direct Submission
SUBMITTED (21-OCT-1996) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jesse.sanger.ac.uk or twinematode.wustl.edu
2 (bases 1 to 18632)
REFERENCE Wilson,R., Almscough,R., Anderson,K., Baynes,C., Berks,M.,
Bontfeld,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones M., Kerhaw,J., Kirsten,J., Laister,N.,
Latteille,P., Lightning,J., Lloyd,C., McMurtry,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Ritten,L., Roopra,A.,
Saunders,D., Showkhen,R., Smaildon,N., Smith,A., Sonhammer,E.,
Stander,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
Wilkinson-Spratt,J. and Woldman,P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
Nature 368 (6466), 32-38 (1994)
94150718
JOURNAL MEDLINE
COMMENT
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
clone C29F3. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
Coding sequences below are predicted from computer analysis, using
the program Genefinder (P. Green, ms in preparation), and other

```


REFERENCE 1 (bases 1 to 115419)
Eukaryotes: Mitochondrial eukaryotes: Metazoa: Chordata: Vertebrata: Eutheria: Primates: Catarrhini: Homiidae: Homo.
AUTHORS Hunt, A.
TITLE Direct Submission
JOURNAL Submitted (06-NOV-1996) Wellcome Trust Genome Camps, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: hummer@anger.ac.uk
COMMENT This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. Unfinished sequence: BK799F10 Contig_ID: 00052 Length: 992 bp Unfinished sequence: BK799F10 Contig_ID: 00061 Length: 885 bp Unfinished sequence: BK799F10 Contig_ID: 00185 Length: 12614 bp Unfinished sequence: BK799F10 Contig_ID: 00949 Length: 830 bp Unfinished sequence: BK799F10 Contig_ID: 00953 Length: 785 bp Unfinished sequence: BK799F10 Contig_ID: 00960 Length: 958 bp Unfinished sequence: BK799F10 Contig_ID: 01506 Length: 6932 bp Unfinished sequence: BK799F10 Contig_ID: 01718 Length: 38510 bp Unfinished sequence: BK799F10 Contig_ID: 02165 Length: 15742 bp Unfinished sequence: BK799F10 Contig_ID: 02195 Length: 2254 bp Unfinished sequence: BK799F10 Contig_ID: 02223 Length: 1000 bp Unfinished sequence: BK799F10 Contig_ID: 02224 Length: 1000 bp Unfinished sequence: BK799F10 Contig_ID: 02225 Length: 1341 bp Unfinished sequence: BK799F10 Contig_ID: 02226 Length: 1016 bp Unfinished sequence: BK799F10 Contig_ID: 02214 Length: 19570 bp.

*** WARNING: Phase 1 High Throughput Genome Sequence ***

* This sequence is unfinished. When sequencing is complete, *
* the sequence data presented in this record will be replaced *
* by a single finished sequence with the same accession number. *
FEATURES
source
1..115419
/organism="Homo sapiens"
/clone="799F10"
/chromosome="22"
BASE COUNT 24169 a 28425 c 26764 g 23990 t 12071 others
ORIGIN
Query Match 1.68; Score 24; DB 34; Length 115419;
st Local Similarity 80.0%; Pred. No. 1.38e+00;
atches 32; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Db 109739 tcttctgttctgttcttcttcacactctgcagctctgaag 109778
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 833 TCTTTGTTTGATGTTTCCATCACTTCAGCAGCTGGAAG 794
RESULT 14
LOCUS D90904 150894 bp DNA BCT 20-NOV-1996
DEFINITION Synchocystis sp. PCC6803 complete genome, 6/27, 630555-781448.
ACCESSION D90904
NID g1652225
KEYWORDS formaldehyde dehydrogenase; ABC transporter; DNA ligase; DNA polymerase III alpha subunit; Gumb protein; LysR transcriptional regulator; Mg-protoporphyrin IX; N-acetylornithine aminotransferase; NADH dehydrogenase subunit 4; NADH dehydrogenase subunit 5; acetate kinase; aspartoacylase(ASP); cation-transporting ATPase; cell division protein FtsH; chemotaxis protein CheB; cytochrome oxidase d subunit I; cytochrome oxidase d subunit II; dieneolactone hydrolase; dihydroflavonol 4-reductase; dihydropterocate pyrophosphorylase; esterase; ferrous iron transport protein B; fibrillin; formaldehyde dehydrogenase (glutathione); high light-inducible protein; leader peptidase I; lysostaphin; malonyl coenzyme A-acyl carrier protein transacylase; methyl-accepting chemotaxis protein (MCP); oligopeptide transport ATP-binding

protein Opp; oxygen independent coporphorphyrinogen III oxidase; penicillin-binding protein 1b; phenoxymethylate dioxygenase; phycoerythrin b subunit; phycoerythrin associated linker protein; phycoerythrin b subunit; potassium channel; protein conferring resistance to acetazolamide, zam; regulatory component of sensory transduction system; regulatory components of sensory transduction system; ribonuclease II; sensory transduction histidine kinase; serine esterase; serine/threonine protein kinase; seryl-L-lysine synthetase; sporulation protein SpoIID; sulfolipid biosynthesis protein SdgB; tRNA-Arg; tRNA-Leu; tyrosyl tRNA synthetase. Synchocystis sp. (strain:PCC6803) DNA.
SOURCE
ORGANISM Synchocystis sp.
REFERENCE 1 (bases 1 to 150894)
Eukacteria: Cyanobacteria: Chroococcales: Synchocystis.
AUTHORS Tabata, S.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-1996) to the DDBJ/EMBL/GenBank databases. Satoshii Tabata, Kazusa DNA Research Institute, Laboratory of Gene Structure 2, 1523-3, Tananuchino, Kisarazu, Chiba 292, Japan (E-mail: tabata@kazusa.or.jp, Tel: +81-438-52-3933, Fax: +81-438-52-3934)
2 (sites)
REFERENCE
AUTHORS Kaneo, T., Sato, S., Kotani, H., Tanaka, A., Asanizu, E., Nakamura, Y., Miyajima, N., Hirosewa, M., Sugiyura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. and Tabata, S.
TITLE Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions
JOURNAL DNA Res. 3 (3), 109-136 (1996)
COMMENT
MEDLINE
Potential protein coding regions were assigned on the basis of similarity search of the ORFs and Genemark analysis.
FEATURES
source
1..150894
/organism="Synchocystis sp."
/strain="PCC6803"
/complement(1..1038)
/note="ORF_ID:s110998"
/codon_start=1
/product="LysR transcriptional regulator"
/db_xref="PID:g1652226"
/transl_table=1
/transl_table=1
/translation="MSDIPFTDQLRIKKAISEGSEFRADRLTYVSOPLVONLU EKOLSVPLFDRGGRKAQLTGHLNLYGKIIITLCOTCRALDNLNGGLIYVA SOTGTYLPLPMIGMRQOQYDPYVOQVSHSTRTAGVANGVDLAIIGSEVPAEIO ETIVLVLPYAEDELALIPVLPPLAOETIOEDLYKLFISLDSOSTIRKVIDVLSO GEIDTRKRIKEMELNSIEAIKNAVOSGLGAFTVTAIEKLENNVLAHPIKRVETIR VLSVIINPNRYRSKASNAFIREVLPOFSTHPDALDEPRFLFANPYSNNNGRDGKD GKSIEIDSV"
complement(1342..1539)
/note="ORF_ID:s11911"
/codon_start=1
/product="hypothetical protein"
/db_xref="PID:g1652227"
/transl_table=1
/translation="MSTGOQARALMBRHQFIKRRQOSMLSRAAEIGVENDFWT VQGRPOSSFRRTYRSNALS"
1659..2603
/note="ORF_ID:s1r1019"
/codon_start=1
/product="hypothetical protein"
/db_xref="PID:g1652228"
/transl_table=1
/translation="MENSVPYRMYGFTLDVFTDQLFGGNPLAVPDAEGLDGOQO KTAIEINISERVFLPVPTETGNGRLRITPKRLDAGHRTTGLTLLGLPPSP LVTIWOEEEPVGLPVTLHYKOGOLVOTETVLAOLDPAKSGSCEDLALLGLSIO LQOEYVPOASCSGLPFLIPLINEALNRSFSPVWOMLNGOMADYCYCLAPGP SLSGSDPKILHGRFAPGLGIAEDPATGSAALGGVLDLRDTPGSHHNGIIOGRAL GRPQDLOTVYKQDGGIRAKVAGRSVLVSEGFNNLGN"
2750..2830
/product="tRNA-Leu"

CDS
complement(3235..4326)
/note="ORF_ID:s110997"
/codon_start=1
/product="hypothetical protein"
/db_xref="pid:g1652229"
/transl_table=11
/translation="MAPKSCNCGDNKQNSNIGEMIVAPYQSFHIGLLALASVW
PLSACIATPSPPAVEONQOVNGROEOALQNSDITYFEETFDPSDEPKRIVGKRG
DYSACITRSROEKHDESEMFPRVYTKFKNKIYATIEGSESPRIALLDTMTKDNPY
PAVEATITGGHNCNEVKVFTSNQSGSDMSYRDGFTNGGPHFEDLNNGWELVYH
VDNRFLYSSTAGSAAPQAQIALQNGVVDVSEPSFTQIHRKASQENDELEIVAW
QDSEKNGFLAAVYANKRALIGELDEGMQMLKXYRDESDMGTLNCLYEDQSNCLNEVY
YDSYPDALRAFLVEAGYIEAKFEAEENNTL"
complement(4338..5693)
/note="ORF_ID:s110996"
/codon_start=1
/product="hypothetical protein"
/db_xref="pid:g1652230"
/transl_table=11
/translation="MNVSHRRYHITTFGCGMKNKADSEEMAGILENTGYNDPNOAD
ILVYNTGSRDQAEQKVSYVLGRQAKRQVPELILVYAGVAQOEGQILRRPDELLE
LYMGQPHNRNLDOLLEQVWAGSOVYATSTELTHMEDITPRESTYSAMVNTYCCNR
CSYCYVNPVRGYEVSORTPEBATIGEMEYVAQSGFKVYTLILGQNDIAGDLDGTTPSGR
HLHTITLDLYHVDLEGIDRLKFPATSHRYFTFRFYEQACQELPKVCEFHFIIPQSGDN
DLKMKAPYTREREKYLQIIEKIRYMPDPAASADVIYVFPETQAEQENNTNLIEVQ
PDLMTAAVSPRGTPTPAAFWMDQLEEVKGRGLNGLVSTQAMERSQRLGVEEVEE
LYVEGENLKSQGVQMGRTGRNRLTFPGQISELLGKTVPVKILTEARAFSLTGEALSVYV
A"
5818..6969
/gene="sqdB"
/note="ORF_ID:s1r1020"
/codon_start=1
/product="sulfolipid biosynthesis protein sqdB"
/db_xref="pid:g1652231"
/transl_table=11
/translation="MRALVYGGDGYCGNATALYSNKREYEVGIGLDSLYRRWMDQLGN
ETLIYIAIRQLDNRWLETGTGKIDLFIGDINDIFPFLINALRQOPDPAVYFGRQSR
PFSMIDRHAVALTQANNVNLGNLNLAYKLEDPCHLVKLTQMTGEYGTNPNDIEEGYT
TIEHGRKQDLEPYKQSGSFYHLSKRVHDSHNHFACKTQWGRADLNGIYGVLTBE
TGMDDLINRLDYDGVFTALNRFICIQALIGHPLVYVKGQOTRGILDIPVTCIELE
AIANPADQGVFNQYETLEFSDVLAQWOKAGADIGLKVIEDHLENPVRELBHFV
NAVNNTLIDIGQPHFLSDLSLNLNATKYKDRVDDKHLPLVYTWNG"
7112..7402
/note="ORF_ID:s1r1698"
/codon_start=1
/product="hypothetical protein"
/db_xref="pid:g1652232"
/transl_table=11
/translation="MADPLTPAISDRICKHNEDHSAIALYAVFGQOTVYTAQNO
AIDPGRMLVYSESGSKTIRIEFPQPLKDESDAHQVILIAKQARSVGKNS"
complement(7439..8134)
/note="ORF_ID:s110995"
/codon_start=1
/product="hypothetical protein"
/db_xref="pid:g1652233"
/transl_table=11
/translation="MTVALDREITIPPSDQGPMDANTQEFMYULKENLECLTAHNP
DVFWGLDLMPVEBGHPETIRVAPQDVVALGPRKGRGSGYRQWQENNOAPQVFEIIS
GNTLEEMTKKAKFYDHHGVGEYVYVDPDNLDTGQRIGELTIIEENAHVVSLLIIG
KFELSAETLIRVYVDPGRPELSTVALATAQEOASCRANEAQRAEREREKLRALAEAEEND
RKALAEAGIDV"
complement(8220..9029)
/note="ORF_ID:s110994"
/codon_start=1
/product="hypothetical protein"
/db_xref="pid:g1652234"
/transl_table=11
/translation="MPTNFFVPSIOMGYVCOKTVISIRNTVIMLHLANTLYWLLG
IGILIGLMTISDGDDELDVEMSLASDVDTIYDVTDELQMGMEGAEVPAALQV
LSFELQVAMITLGLDIPSLMGVGLVNTLVAVTNTPELMKAGAGITFVSLAH
SIMDLRLSRPAHLKFTFSQDVSTERYIGCTGYTSKKPLVLANGTIGQAHVADIVS
NLTITSVSLPDVATYIPAHNEILLIDOSPFGYGLATAKDSYSDCKWLKS"
complement(8962..110059)
CDS
complement(3235..4326)
/note="ORF_ID:s110997"
/codon_start=1
/product="hypothetical protein"
/db_xref="pid:g1652229"
/transl_table=11
/translation="MAPKSCNCGDNKQNSNIGEMIVAPYQSFHIGLLALASVW
PLSACIATPSPPAVEONQOVNGROEOALQNSDITYFEETFDPSDEPKRIVGKRG
DYSACITRSROEKHDESEMFPRVYTKFKNKIYATIEGSESPRIALLDTMTKDNPY
PAVEATITGGHNCNEVKVFTSNQSGSDMSYRDGFTNGGPHFEDLNNGWELVYH
VDNRFLYSSTAGSAAPQAQIALQNGVVDVSEPSFTQIHRKASQENDELEIVAW
QDSEKNGFLAAVYANKRALIGELDEGMQMLKXYRDESDMGTLNCLYEDQSNCLNEVY
YDSYPDALRAFLVEAGYIEAKFEAEENNTL"
complement(4338..5693)
/note="ORF_ID:s110996"
/codon_start=1
/product="hypothetical protein"
/db_xref="pid:g1652230"
/transl_table=11
/translation="MNVSHRRYHITTFGCGMKNKADSEEMAGILENTGYNDPNOAD
ILVYNTGSRDQAEQKVSYVLGRQAKRQVPELILVYAGVAQOEGQILRRPDELLE
LYMGQPHNRNLDOLLEQVWAGSOVYATSTELTHMEDITPRESTYSAMVNTYCCNR
CSYCYVNPVRGYEVSORTPEBATIGEMEYVAQSGFKVYTLILGQNDIAGDLDGTTPSGR
HLHTITLDLYHVDLEGIDRLKFPATSHRYFTFRFYEQACQELPKVCEFHFIIPQSGDN
DLKMKAPYTREREKYLQIIEKIRYMPDPAASADVIYVFPETQAEQENNTNLIEVQ
PDLMTAAVSPRGTPTPAAFWMDQLEEVKGRGLNGLVSTQAMERSQRLGVEEVEE
LYVEGENLKSQGVQMGRTGRNRLTFPGQISELLGKTVPVKILTEARAFSLTGEALSVYV
A"
5818..6969
/gene="sqdB"
/note="ORF_ID:s1r1020"
/codon_start=1
/product="sulfolipid biosynthesis protein sqdB"
/db_xref="pid:g1652231"
/transl_table=11
/translation="MRALVYGGDGYCGNATALYSNKREYEVGIGLDSLYRRWMDQLGN
ETLIYIAIRQLDNRWLETGTGKIDLFIGDINDIFPFLINALRQOPDPAVYFGRQSR
PFSMIDRHAVALTQANNVNLGNLNLAYKLEDPCHLVKLTQMTGEYGTNPNDIEEGYT
TIEHGRKQDLEPYKQSGSFYHLSKRVHDSHNHFACKTQWGRADLNGIYGVLTBE
TGMDDLINRLDYDGVFTALNRFICIQALIGHPLVYVKGQOTRGILDIPVTCIELE
AIANPADQGVFNQYETLEFSDVLAQWOKAGADIGLKVIEDHLENPVRELBHFV
NAVNNTLIDIGQPHFLSDLSLNLNATKYKDRVDDKHLPLVYTWNG"
7112..7402
/note="ORF_ID:s1r1698"
/codon_start=1
/product="hypothetical protein"
/db_xref="pid:g1652232"
/transl_table=11
/translation="MADPLTPAISDRICKHNEDHSAIALYAVFGQOTVYTAQNO
AIDPGRMLVYSESGSKTIRIEFPQPLKDESDAHQVILIAKQARSVGKNS"
complement(7439..8134)
/note="ORF_ID:s110995"
/codon_start=1
/product="hypothetical protein"
/db_xref="pid:g1652233"
/transl_table=11
/translation="MTVALDREITIPPSDQGPMDANTQEFMYULKENLECLTAHNP
DVFWGLDLMPVEBGHPETIRVAPQDVVALGPRKGRGSGYRQWQENNOAPQVFEIIS
GNTLEEMTKKAKFYDHHGVGEYVYVDPDNLDTGQRIGELTIIEENAHVVSLLIIG
KFELSAETLIRVYVDPGRPELSTVALATAQEOASCRANEAQRAEREREKLRALAEAEEND
RKALAEAGIDV"
complement(8220..9029)
/note="ORF_ID:s110994"
/codon_start=1
/product="hypothetical protein"
/db_xref="pid:g1652234"
/transl_table=11
/translation="MPTNFFVPSIOMGYVCOKTVISIRNTVIMLHLANTLYWLLG
IGILIGLMTISDGDDELDVEMSLASDVDTIYDVTDELQMGMEGAEVPAALQV
LSFELQVAMITLGLDIPSLMGVGLVNTLVAVTNTPELMKAGAGITFVSLAH
SIMDLRLSRPAHLKFTFSQDVSTERYIGCTGYTSKKPLVLANGTIGQAHVADIVS
NLTITSVSLPDVATYIPAHNEILLIDOSPFGYGLATAKDSYSDCKWLKS"
complement(8962..110059)
CDS
complement(3235..4326)
/note="ORF_ID:s110997"
/codon_start=1
/product="hypothetical protein"
/db_xref="pid:g1652229"
/transl_table=11
/translation="MAPKSCNCGDNKQNSNIGEMIVAPYQSFHIGLLALASVW
PLSACIATPSPPAVEONQOVNGROEOALQNSDITYFEETFDPSDEPKRIVGKRG
DYSACITRSROEKHDESEMFPRVYTKFKNKIYATIEGSESPRIALLDTMTKDNPY
PAVEATITGGHNCNEVKVFTSNQSGSDMSYRDGFTNGGPHFEDLNNGWELVYH
VDNRFLYSSTAGSAAPQAQIALQNGVVDVSEPSFTQIHRKASQENDELEIVAW
QDSEKNGFLAAVYANKRALIGELDEGMQMLKXYRDESDMGTLNCLYEDQSNCLNEVY
YDSYPDALRAFLVEAGYIEAKFEAEENNTL"
complement(4338..5693)
/note="ORF_ID:s110996"
/codon_start=1
/product="hypothetical protein"
/db_xref="pid:g1652230"
/transl_table=11
/translation="MNVSHRRYHITTFGCGMKNKADSEEMAGILENTGYNDPNOAD
ILVYNTGSRDQAEQKVSYVLGRQAKRQVPELILVYAGVAQOEGQILRRPDELLE
LYMGQPHNRNLDOLLEQVWAGSOVYATSTELTHMEDITPRESTYSAMVNTYCCNR
CSYCYVNPVRGYEVSORTPEBATIGEMEYVAQSGFKVYTLILGQNDIAGDLDGTTPSGR
HLHTITLDLYHVDLEGIDRLKFPATSHRYFTFRFYEQACQELPKVCEFHFIIPQSGDN
DLKMKAPYTREREKYLQIIEKIRYMPDPAASADVIYVFPETQAEQENNTNLIEVQ
PDLMTAAVSPRGTPTPAAFWMDQLEEVKGRGLNGLVSTQAMERSQRLGVEEVEE
LYVEGENLKSQGVQMGRTGRNRLTFPGQISELLGKTVPVKILTEARAFSLTGEALSVYV
A"
5818..6969
/gene="sqdB"
/note="ORF_ID:s1r1020"
/codon_start=1
/product="sulfolipid biosynthesis protein sqdB"
/db_xref="pid:g1652231"
/transl_table=11
/translation="MRALVYGGDGYCGNATALYSNKREYEVGIGLDSLYRRWMDQLGN
ETLIYIAIRQLDNRWLETGTGKIDLFIGDINDIFPFLINALRQOPDPAVYFGRQSR
PFSMIDRHAVALTQANNVNLGNLNLAYKLEDPCHLVKLTQMTGEYGTNPNDIEEGYT
TIEHGRKQDLEPYKQSGSFYHLSKRVHDSHNHFACKTQWGRADLNGIYGVLTBE
TGMDDLINRLDYDGVFTALNRFICIQALIGHPLVYVKGQOTRGILDIPVTCIELE
AIANPADQGVFNQYETLEFSDVLAQWOKAGADIGLKVIEDHLENPVRELBHFV
NAVNNTLIDIGQPHFLSDLSLNLNATKYKDRVDDKHLPLVYTWNG"
7112..74

```

/note="ORF_ID:sl10993"
/codon_start=1
/product="potassium channel"
/db_xref="PID:g1652235"
/transl_table=11
/translacion="MGLSSSSQEOENLNLIDROGRRLRQELMAGAITLAGLFTVGTAWY
RYEDWTMLDAFWITTLTATVGGEGHNPSPASRETTILLGLGLITGMWRPFE
AFIIOGYRODSLRRROEKQVIERLADHYITLIGCGTGOOIAFEFVENIPVUIDASPD
VITIOAKRDRAVIUGDATTDEILLAHIERAICVLSLSDAENLYVLSKTLNPKRI
FALAPASSEEAAYGKLKRAGADEVAVSPITITGKRLAAALAPQVVSFVDGITGADRSEI
YMEFIRGAGDCPYITGOTLREAO LRAGSGALITLAIIRRODKRLIVGPGDTHLLDADSL
ICGTEGRLRALNQLCLPLNPARVRLPKNR"
10183..11472
/gene="argD"
/note="ORF_ID:sl11022"
/codon_start=1
/product="N-acetylornithine aminotransferase"
/db_xref="PID:g1652236"
/transl_table=11
/translacion="MTKSPVSEVYEAQAFVTDLSPAEFRTADFDITVNMVGRPPID
AIRGSGSTLMDTEGKSYLDVFAVGIACTGHAHPALVRAVSDIOKLHHVSNLYTIR
EGEELAKWIIYHSCADRVFCNSGAEENEAIKLYRRTAVTDLEQPIYILTKASF
HGHTLATITATGQPKIQTIDPLVPGFDIYPINDIRSLKNKVAIDLSDGNSVAIFLE
PLOGEGVRPDDIADLFKRVREIDQNDLILVFEVOVGVRTADLWGEYHLEGPDIFFE
TSARKGLAGVPIGAMMKKFCVDVEFEPMNHAISTGGNLACAGAAVATKTEGDRLLDN
VQARGEOLRSGLAIEKNOYPTLFEVGMQDINGLETSAESSLTSVEIVKAMOGGLL
LAPAGPVLRFVPLVYVTEAIEAQAVETLMOATVATV"
11599..11913
/note="ORF_ID:sl11023"
/codon_start=1
/product="hypothetical protein"
/db_xref="PID:g1652237"
/transl_table=11
/translacion="MKMLQMLPRLPPQEOFDLLENKTLIERIISTGOTIPGGEWY
DQBRPEVNVLLQCKAETTYDTGAVISLGGODYLLIPRQKHRTYFTTHEPPCIYLAHL
FS"
11920..12561
CDS
Note: remainder of annotations omitted.
Query Match 1.6%; Score 24; DB 20; Length 150894;
Best Local Similarity 81.6%; Pred. No. 1.38e+00;
Matches 51; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Db 4235 aaactctgatatggaccatgactatcaccctcccaaa 4272
||||| ||||| ||||| ||||| ||||| |||||
OY 1388 AAACATATGATGTGGAGAAGAGCACTACATCTCCGCCAA 1425

RESULT 15
LOCUS ECACERA 1344 bp DNA BCT 12-SEP-1993
DEFINITION E. coli acea gene coding for isocitrate lyase (EC 4.1.3.1).
ACCESSION X07543
NID 940884
KEYWORDS acea gene; isocitrate lyase; isocitratase.
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 1344)
Cozzone,A.J.
Submitted (06-MAY-1988) Cozzone A.J., University of Lyon,
Laboratory of Molecular Biology, 43 Boulevard du onze Novembre,
69622 Villeurbanne, France
2 (bases 1 to 1344)
Reul,C., Biecher,F., Duclos,B., Cortay,J.C. and Cozzone,A.J.
Nucleotide sequence of the acea gene coding for isocitrate lyase in
Escherichia coli
Nucleic Acids Res. 16 (12), 5689 (1988)
86262573
JOURNAL MEDLINE
FEATURES
Source Location/Qualifiers
1..1344

```


/organism="Escherichia coli"
/strain="K12"
/map="91 min"
27..30
/note="pot. RNA binding site"
40..1344
/note="isocitrate lyase (AA 1-434)"
/codon_start=1
/db_xref="PID:g40885"
/db_xref="SWISS-PROT:P05313"
/translation="MKTRQOIELOKEMTOPRMGRTPRYSADYVYKLRGSVNPECT
LALGAAKMRRLHGESKGYINSLGALTGGAALQQAAGIEAVYILSGMVAADAMWR
PACIRISRSIROTRCEPAVERINNTERRADQIQWSAGIEGDPYVDFLPIVADAEA
GFGVNAFELMKAMIEAGAAVHFEDOLASVKCGHMGKVLPTQEAIOKLVPARL
AAVTVPTLLVARTDADADLITSDCDPYDSEFTIGERTSEGFRTHAGIEQALSRG
LAVAPYADLVWCETSTPDLELARFPAQAIHAKYPGKLIAYNCSFSFMQKNLDDKTI
SFGEOISDMGVKROFTTLAGIHSMMFMEDLANAYNOGGMKHYEKKVQPEPAAKD
GITFVSHQDVGIGTGDKYTTTIQGTSSVTALTGSTESQF"

COUNT 321 a 353 c 386 g 284 t
-GIN

Query Match 1.5%; Score 23; DB 20; Length 1344;
Best Local Similarity 81.1%; Pred. No. 5.59e+00;
Matches 30; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 983 ctataactgctgcgcgctcctcaactgcagaaaaa 1019
||||| ||||| || || ||||| |||||
QY 1488 CTTATTACTGCTGCACATAATCACTGAAGAAAAA 1524

Search completed: Thu Aug 21 10:01:07 1997
Job time : 2597 secs.